Computational Development of the Nanoporous Materials Genome

Peter G. Boyd†, Yongjin Lee†, Berend Smit*†‡

†Laboratory of Molecular Simulation, Institut des Sciences et Ingénierie Chimiques, École Polytechnique Fédérale de Lausanne (EPFL), Rue de l'Industrie 17, CH-1951 Sion, Valais, Switzerland

‡Department of Chemical and Biomolecular Engineering, University of California, Berkeley 94720, USA

E-mail: berend.smit@epfl.ch

Phone: +41 21 693 0079

H1 Abstract.

There is currently a large push towards big data and data mining in materials research to accelerate discovery. Zeolites, metal-organic frameworks (MOFs) and other related crystalline porous materials are not immune to this recent phenomenon, as evidenced by the proliferation of porous structure databases and computational gas adsorption screening studies over the past decade. The strive to identify the best materials for a variety of gas separation and storage applications has not only led to collections of thousands of synthesised structures, but the development of hypothetical material building algorithms. The materials databases assembled with these algorithms expand greatly on the range of complex pore structures that have been synthesised, with the rationale that we have discovered only a small fraction of realisable structures and expanding upon these will accelerate rational design. In this review, we highlight some of the methods developed to build these databases, and some of the important outcomes resulting from large-scale computational screening efforts.

Summary

In the field of nanoporous materials discovery, we are witnessing the emergence of big data analytics combined with traditional computational thermodynamics calculations. This review turns a critical eye

on the current state of the art, with a focus on computational database generation and results from large-scale screening for gas separations.

H1 Introduction.

The discovery, in the late 19th century, that zeolites could trap interesting and valuable particles in their pores opened up an entire field of research[1,2]. This seemingly simple phenomenon was an enormous boon to the oil and gas industry, seeing as how cheap, but effective porous materials could serve as a catalyst for hydrocarbon cracking. From their humble beginnings (being carved out of rock faces), zeolites, and their ability to selectively trap guests in their pores, have become integral in not only the oil and gas industry, but in detergents (as ion exchangers) and natural gas purification to name a few[1]. Zeolites have since enjoyed an almost exclusive dominance in porous materials research until the late 20th century, when we began to observe the creation of more diverse materials in terms of chemistry, network connectivity, and physical properties.

At present, there are a little over 200 known zeolite structures, which is only a small fraction of the total number of structures that have been predicted[3]. In addition, if we were also able to expand the chemical diversity of these materials beyond the conventional Si^{4+}, O^{2-}, and Al^{3+} ions found in zeolites, one could envision designing materials for virtually any gas separation application. Thus when the first articles arose in the 1990's characterising Metal-Organic Frameworks (MOFs)[4–8], and later Covalent Organic Frameworks (COFs)[9,10], Zeolitic Imidazolate Frameworks (ZIFs)[11], and Porous Polymer Networks (PPNs)[12] the excitement was palpable. It was recognised early-on by Yaghi, O'Keeffe, and coworkers[13] that the increase in potential diversity posed by these novel porous materials comes with completely new challenge; finding the best material in a continuously increasing nanoporous haystack. This is a seemingly unbounded problem, as one can envision millions of potential combinations of chemical species to form

novel MOFs. This introduces a significant uncertainty on the form of the resulting coordination polymer, an issue that can have a dramatic effect on material properties and performance. However, by observing common trends in linker-metal coordination geometries, a concept of rational design called reticular chemistry was introduced[14]. Here, one can have a reasonable guess of the resulting material form by abstracting coordination compounds to discrete building blocks called Structural, or Secondary, Building Units (SBUs) and predicting the most probable (read symmetric) arrangement of these SBUs in a 3D network. This method, which applies concepts from graph and tiling theory[15–17], has guided a significant number of researchers in their synthetic design approaches and can be nicely summarised by the various reviews put forth by Yaghi & O'Keeffe[18–20], and topology-based studies of Blatov & Proserpio[21,22]. What it does not do, however, is give researchers an idea on how well materials will perform for a given application. Thus, while reticular chemistry has enabled researchers to conceive of, and successfully synthesize novel structures with desired connectivity, there is still an enormous cost to performing these steps only to find a rationally designed MOF doesn't adsorb much of a desired gas species (e.g. CO_2 for clean energy applications). This raises a related question; how many applications have each synthesised material been studied for? In other words, how many properties have been overlooked for a given material due to limits on time, cost and research ability?

This is where computational researchers have become integral in the development of rational design strategies. Even in an optimal setting, synthesizing, characterizing, and testing a new MOF can take many months. Hence, we can only expect to synthesize a small fraction of the millions of possible MOFs. As predicting a MOFs gas adsorption behaviour on a computer is significantly faster (and cheaper) than performing the analogous experiments, computational research is playing an increasingly important role in developing strategies to identify the most promising materials for a desired application. The computational challenge here is to ensure one uses models that provide a reasonable view of the

thermodynamic and transport properties exhibited by a material (in some cases even before synthesis![23]). Moreover, accurate computational models can provide detailed information about gas-framework interactions resolved to the atomic level, providing an unprecedented level of molecular information supporting the design of next generation materials[24,25].

The idea of augmenting discovery programs with computation is not new. For example, computational screening is frequently used in the pharmaceutical industry as an initial step to identify molecules as potential drug candidates[26]. In this case, virtual molecule libraries are screened for potential candidates computationally, which are then subject to experimental clinical trials to evaluate them *in vivo*. In addition, trends in activity with respect to a drug target are typically identified by abstracting molecular properties in the spatial domain. Here, computational discovery is typically the first step, always to be followed by experimental work. It would seem obvious to adopt a similar strategy for nanoporous materials design, and sure enough in the past 5 years we have witnessed the proliferation of studies on virtual high-throughput screening of porous materials for gas capture and storage applications[27–37].

One major challenge in this field is that, unlike the many organic molecule databases available for pharmaceutical screening, there was simply not enough available data to perform large-scale materials screening studies. The default study involved the screening 10 to 20 materials at a time[38] in order to tease out typical 1-dimensional structure-property relationships. As we will show in this review, only in very rare occasions the optimal material can be captured with a single volcano plot (a 1-dimensional plot of property vs activity where there is a clear peak related to a direct correlation between structural property and performance); in most cases finding the optimal material is a high-dimensional optimization problem of which the solution looks more like set of chemically disconnected volcanos.

These insights could only be obtained because in the computational domain we have been able to supplement these 10-20 materials with libraries of thousands of computationally generated (hypothetical) materials. These libraries required novel strategies to assemble hypothetical materials *in silico*, which borrow heavily from the concept of reticular chemistry. The methods, the databases of millions of porous crystalline structures that have arisen, their computed performances for a range of applications, and the insight these studies have provided have been collected in what is called the nanoporous materials genome, and will be the subject of this critical review.

H1 Database Development and the Quest for Diversity.

We start with a discussion on some of the methods used to build databases of nanoporous materials, and how they were used to develop and curate databases for use in high throughput screening studies. It should be mentioned that, while chemically different, all of the different subclasses of nanoporous materials are treated essentially the *same* in what follows. That is to say, the algorithms that assemble these materials, and those which evaluate gas adsorption properties could be applied to MOFs, ZIFs, zeolites, or COFs. How each class of material performs for a particular application is largely dependent on their physical pore characteristics, a simplified measure of which are presented in Figure 1.

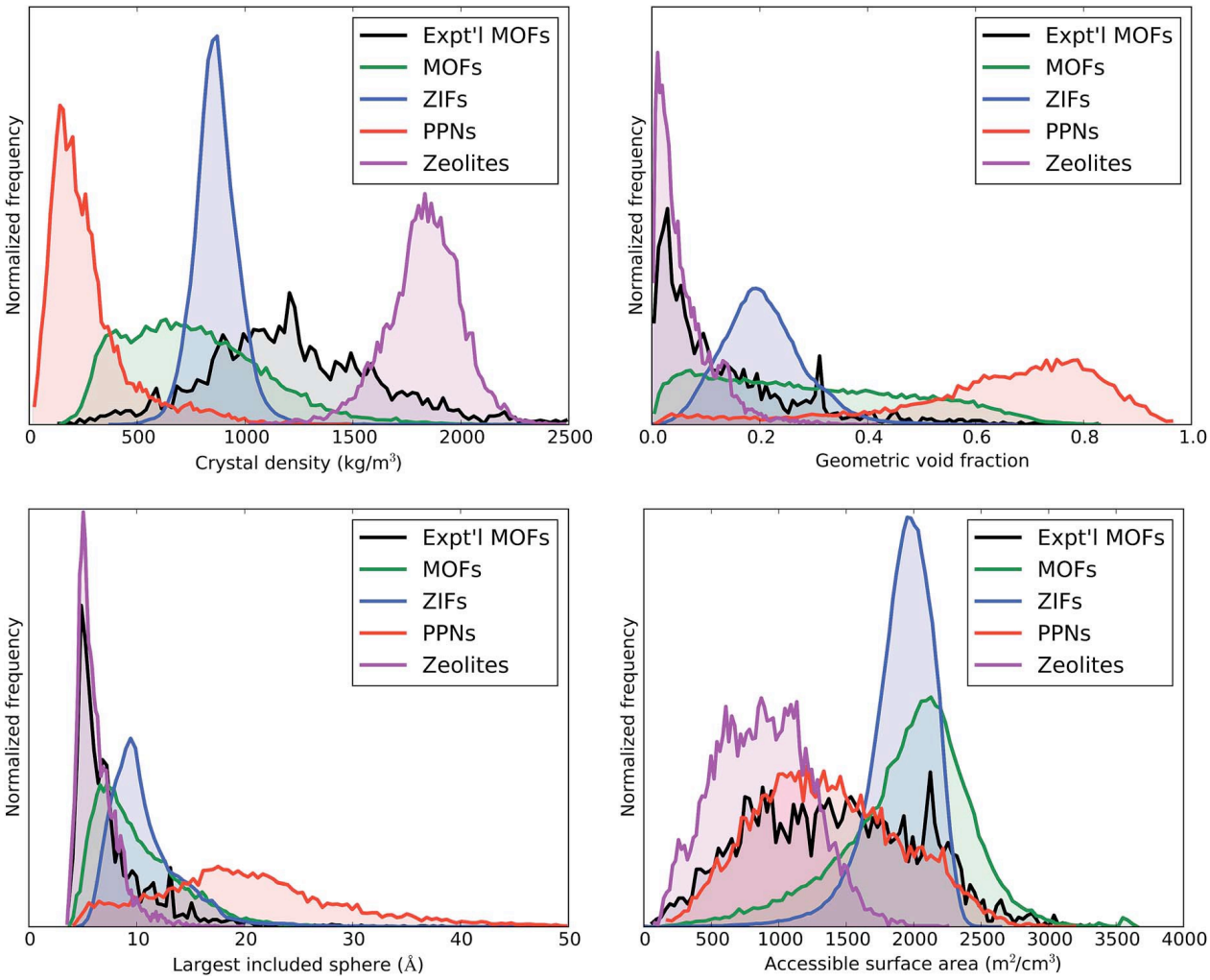

Figure 1: Property distributions of the MOF, zeolite, PPN, and ZIF databases in the nanoporous materials genome. Each structure class possessing a unique distribution of chemistry and (from L-R) crystal density, void fraction, largest included sphere, and surface area. Image reproduced with permission from Ref [30].

With the exception of the Computation REady (CoRE) database[39], which contains synthesised MOF structures from the Cambridge Structural Database (CSD)[40], all of these databases were developed *in silico* with newly developed algorithms. While many articles have discussed the differences between these structure generation methods, we argue that 4/5ths of these algorithms are ultimately sampling a

very similar chemical space only in a slightly different manner, where some methods are more efficient than others.

Table 1: A brief summary of the nanoporous materials databases and the screening studies performed on them.

Database	Size	Topologies[†]	Property Screening Studies
hZeolites	331,172	331,172	CO_2:N_2[41,42], CO_2:CH_4 and CO_2:N_2 membrane separations[43], CH_4[29,30], CO_2:CH_4[44], CO_2:H_2[45], CO_2:H_2O[46], ethane:ethene[47], ethanol:H_2O[48], hydrocarbons[48], Xe:Kr[49], dielectric constants and Bulk moduli[50],
hMOFs[51]	137,953	6 (**pcu**, **sra**, **dia**, **tbo**, **nbo**, **fcu**)[52]	CH_4[30,51], CO_2:N_2[53], Xe:Kr[49,54], CH_4:ammonia:H_2O[55], H_2[56–58], CO_2:H_2[59], CO_2:N_2:CH_4 membrane separations[60], flexibility[61], shape selectivity[62]
PPNs[31]	17,846	1 (**dia**)	CH_4[30,31], Xe:Kr[49]
pcu-hMOFs[33]	116	1 (**pcu**)	CH_4[30,33]
Covalent Organic Polymers (COPs)[63]	50	50	CO_2:N_2[63]
MOF-74-hMOFs[64]	61	1 (**etb**)	CO_2[64]
CoRE[39]	4764	350	CH_4[30,39], C_8 aromatic separation[65], CO_2:N_2:H_2O[66], propylene:propane[67], CO_2:N_2:CH_4[68], Xe:Kr[49,69]
Ref [67] and [69]- Independent mining and cleaning of the CSD, not necessarily the CoRE database.			

Ref [57] – modified database of 18,383 hMOFs with Mg alkoxide functional groups using the same algorithm as ref [51].

† Topologies referenced here (**pcu**, **sra**, etc.) are unique three letter identifiers used to distinguish nets in the Reticular Chemistry Structure Resource (RCSR)[70].

An early example of generating porous materials *in silico* is *a* database of over 2 million hypothetical zeolites[71,72]. These predicted zeolite structures were constructed with a Monte Carlo sampling method, where tetrahedral Si^{4+} ions were placed in a unit cell and temperature controlled adjustments of their symmetry-allowed positions were performed to identify low-energy structures. The database has been screened for a number of applications, including CO_2 capture[41,42] and separation[43], methane storage[30], ethane:ethane separations[47], and ethanol:water separations[48]. Interestingly, some discoveries made in these studies demonstrate properties that have yet to be expressed in synthetically realised structures, such as zeolitic pores of wide diameter, yet containing strong adsorption sites for CO_2, making an ideal setting for strong adsorption and fast diffusion[41].

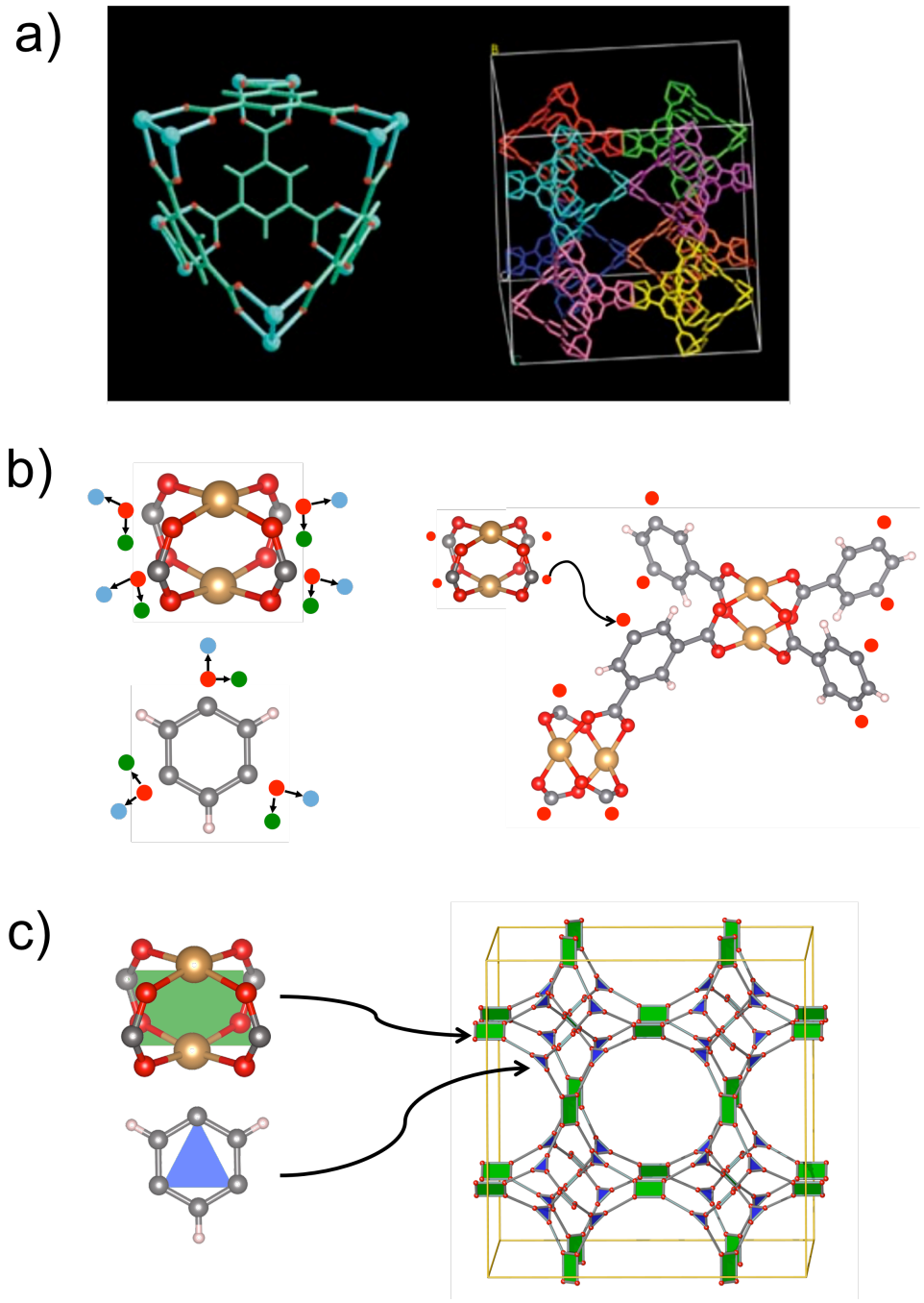

Figure 2: Brief descriptions of building the prototypical MOF, HKUST-1, with the different MOF assembly methods discussed in the text. a) The AASBU method, which is a compute-intensive method, requires creating large hybrid building blocks (shown on left) to efficiently build HKUST-1. The MOF is shown on the right, where each building block is represented in a different color. b) the 'tinker toy' algorithm of Wilmer et al. uses alignment parameters for each SBU shown as the red, green and blue pseudo-atoms on each SBU (left). When assembling HKUST-1, an initial 'seed' SBU is placed, then a recursive search of all possible bonding combinations is attempted until there are no free bonds. The topology of HKUST-1 is encoded in the alignment parameters. Only the red alignment parameters are shown on the right for clarity. c) The topology-based algorithms use known topologies as templates for assembling MOFs. For HKUST-1, the ***tbo*** *topology is used, where the Cu SBU is oriented on the 'square' nodes and the benzene SBU is oriented on the 'triangular' nodes. Legend: Cu, orange; O, red; C, grey; H, pink. Image a) was reproduced with permission from ref* [73].

The first known algorithm for developing MOF structures was presented in 2000[74], the primary purpose of which was to predict new structures rather than construct databases of MOF structures. The method, titled Automatic Assembly of Secondary Building Units (AASBU), borrowed ideas from zeolite and bulk material prediction algorithms[75–77]. Namely, at the core of the AASBU method is a global optimization technique, where MOF building blocks (SBUs) are treated as rigid units containing 'sticky' atoms, and as these building blocks are randomly perturbed in a large simulation box as a function of temperature, nearby sticky atoms will adhere and break to dictate the formation of extended coordination polymers[73]. This algorithm samples a relatively large number of degrees of freedom during the assembly process, which in turn, permits the possibility of building many unique structures from a single selection of SBUs. However, sampling so many degrees of freedom requires the algorithm to carry out an expensive sorting step through large numbers of produced structures to identify unique and crystalline materials, which limits the use of this method to generate large databases. Methods have been presented that reduce this sampling problem, for example by agglomerating together small SBUs to create larger building blocks with fewer sticky atoms[73,78], however this implies that one has a pre-conceived idea of the desired final structure (see Figure 2a) and reduces the predictive power of the algorithm.

The remaining algorithms simplify the process of constructing crystals by making the assumption that SBUs will self-assemble to form pre-defined 3D graph patterns, or nets. The first of said algorithms was published in 2012 along with a database of 138,000 hypothetically generated materials[51]. Here chemical building blocks were extracted from x-ray resolved experimental MOF structures based on a logical partition of the chemical units, such that each building block represented simple geometric shapes (squares, triangles, tetrahedra, octahedra, etc.). Points at which building blocks would form bonds to one another in the experimental structure were parameterised with orthogonal alignment vectors shown in Figure 2b, such that if one were to 'snap' two SBUs together by aligning these vectors, a small fragment

of a crystal was formed. The method of assembly proceeds by growing MOF clusters in this fashion, starting with a single SBU and iteratively attempting to add new SBUs to existing un-bonded SBU sites until either an upper limit of attempts is made, or there are no more bonds to attach new SBUs to (resulting in a new hypothetical MOF). This method was aptly named the tinker-toy algorithm, for its way of snapping together rigid building blocks to form a lattice.

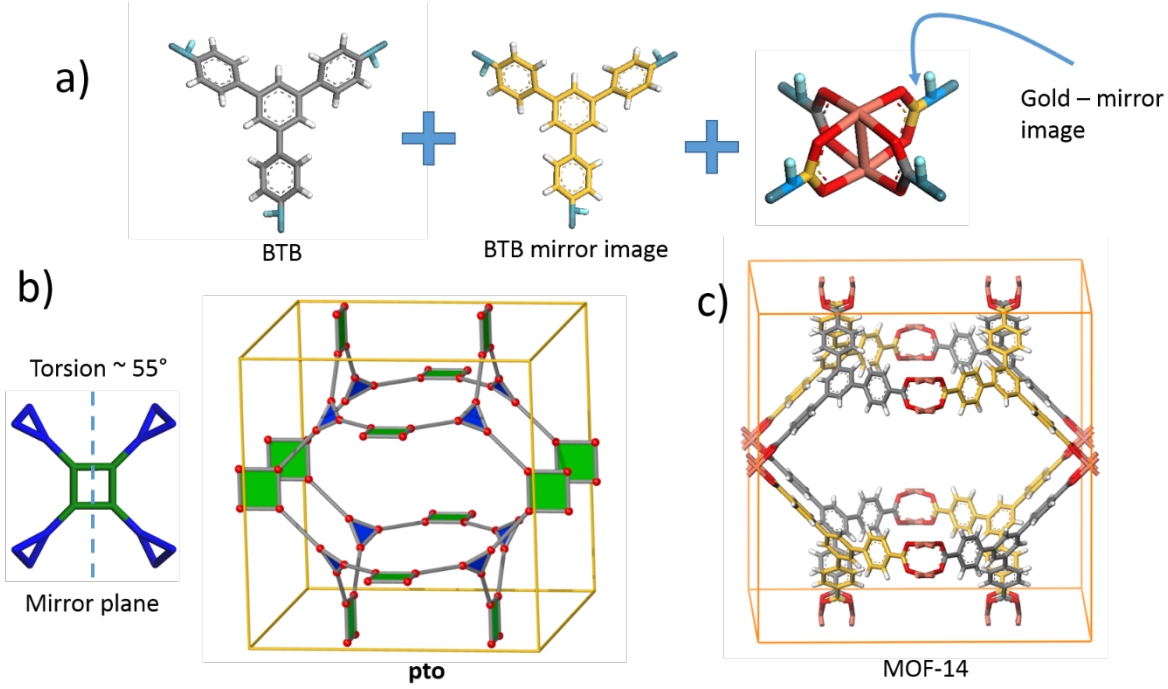

*Figure 3: Challenges associated with assembling new MOFs with the tinker-toy approach. Structural symmetry may require explicit vectors for SBUs and their mirror image. a) The SBUs required to build MOF-14, b) a visual depiction of the **pto** net, where the square planar nodes contain trigonal nodes rotated 55° out of plane. Mirror symmetry in this net yields chirality in the trigonal node coordination geometry. c) MOF-14, where the chiral image of the benzene tri-benzoic acid (BTB) ligand is coloured yellow.*

A particular challenge to using this algorithm is hinted at in its colloquial name, that is to say, the building blocks are destined to snap together in a very specific way, such that one would only be able to construct MOF structures with the same underlying connectivity as the initial experimental structure. Moreover, to increase the diversity of underlying patterns (or nets) will sometimes require a significant increase in the number of alignment vectors, which can be non-obvious. As an example, consider the case of building MOF-14[79]. Figure 3 shows that this MOF can be thought of consisting of two separate building blocks, the

di-copper subunit, commonly known as a paddlewheel, and the benzene tri-benzoic acid (BTB) linker. The symmetry of this network yields two unique BTB groups that are chiral images of one another. Thus, to successfully assemble this material, one would need bond vectors for the copper paddlewheel and *two* separate BTB groups with non-superimposable orientations. In addition, considering the successful assembly of this material would require the proper placement of the chiral images on the paddlewheel, the time required to sample all possible bonding combinations would be non-trivial. The time complexity of sampling the possible permutations of combining nodes in a growing graph is known to be O(n!)[80], where n is the number of SBUs of the growing MOF. The algorithm could therefore take *years* to find the correct combination of SBUs. The limitation of the tinker-toy algorithm can be best illustrated by a topological analysis of the resulting network topologies, which showed the 138,000 structures were made with only 6 underlying network topologies, most of which were primitive cubic (**pcu**)[52]. As a reference, a 2011 topological study on coordination polymers found in the Cambridge Structural Database showed 4709 structures could be found in more than 20 unique topologies (the largest portion of these were, in fact, **pcu** at 9.2%)[22]. Likewise it was shown that the CoRE database of experimental structures contains more than 350 unique topologies (again **pcu** ranking first with 16% of the structures)[39]. It was recently shown that different topologies can yield different minimum and maximum values for surface area and void volume, and the geometries of these structures have an implicit effect on performance[81–83], particularly when considering properties that are more sensitive to geometry and chemistry, such as CO_2 adsorption at low partial pressures. Thus, in general, it would be advantageous to increase the range of topologies included in a database of materials. These observations motivated the development of more efficient approaches discussed below[84,85].

The following algorithms sample the same search space as the tinker-toy approach, however reduce the dependence of computational material assembly on SBU alignment parameters, and instead attempt to best-fit SBUs onto pre-defined net templates. It should be stressed that while recent articles[84] and

reviews[86,87] have distinguished these assembly algorithms as being either “bottom-up” or “top-down” when addressing the tinker-toy and topology-based algorithms, they are essentially performing the same task, though the latter are arguably much more efficient at doing so.

The first topology-based algorithm for generating porous materials was presented in 2014[88] and is included in the porous characterisation software, Zeo++[89,90]. Here, each SBU is identified by their bonding connection sites, such that the algorithm could abstract a certain shape and coordination number from the molecule. We stress that these connection sites do not guide the formation of a specific topology, only identify how the SBU should be oriented in the provided template. The template itself is a three-dimensional net, where the placement of nodes, edges and unit cell dimensions are pre-defined, and obtained from the Reticular Chemistry Structure Resource (RCSR) [70]. To produce structures, the nets cell dimensions are adjusted to best fit the desired SBU geometries. In most cases, this can provide valid structures however, as the geometry of the SBUs deviate farther from the net node geometries, it is possible to observe atomic collisions. Collisions can be dealt with by either discarding the structure, or by energy relaxations. With regards to energetic relaxations, a good first pass is to use a classical force field with the exceptional condition that *all* of the bonding information is known prior to structure assembly. In other words, the intra-SBU bonding is tabulated prior to building structures, and bonds are drawn between connecting atoms of two adjacent SBUs during assembly. This avoids ambiguity in the desired bonding of a structure when the structural minimization takes place. Without this information, artificially close atoms could mistakenly be considered bonded together resulting in odd structural motifs and convergence problems during optimization.

Because this algorithm uses nets to guide the generation of materials, issues such as the presence of coordination chirality that arise from the tinker-toy algorithm are non-existent. One simply needs a

single representative SBU to orient on all the nodes in a net, as the net inherently expresses the chirality in its nodes. It is worthy of note that this algorithm (or earlier versions of it) have been used to generate several hypothetical databases of porous materials with a record number of topologies[31,33,63,83,91]. While the sum of these structures consist of roughly 50 topologies, the RCSR contains a total of 2719 nets as of this writing, leaving a large amount of room for further structural diversity[70].

Shortly following this work, the AuToGraFS (Automatic Topological Generator for Framework Structures) algorithm[92] was presented, which contains many of the same design details as the method in Zeo++. Of notable difference is the streamlined integration of the code with a molecular dynamics program, GULP (General Utility Lattice Program)[93]. The authors have anticipated the need for structural relaxation of generated hypothetical materials, and have taken steps to include atomistic bonding and tailored force field parameters in their structure generation protocol[94,95] to ensure that the resulting structures are in a reasonable configuration. While the program appears ideal for the generation of a large database of diverse materials, to date no database has been reported using this code.

More recently, the algorithm ToBasCCo (Topology Based Crystal Constructor) demonstrated that *in silico* assembly of these structures could be performed using underlying graphs, a level of abstraction from the 3-dimensional net[96]. The underlying principle is the same as AuToGRaFS[92] and Zeo++[88–90] in that a net serves as a template for hypothetical material assembly. However, they demonstrate that geometric distortions arising from chemical bonding within the SBUs can be encoded into the edges and nodes of a net prior to assembly, such that one can obtain a 'best fit' of the SBUs to the desired net, reducing the probability of significant atomic overlap.

A key question about these databases is whether a randomly selected structure can be synthesised[41,88]. Most databases use building blocks that are known chemicals, which limits the chemical diversity but does remove some uncertainty about synthesizability. In addition, most databases have used some energetic indicator to ensure that, from a thermodynamic point of view, the structures are sufficiently stable. However, our knowledge about the conditions and why a particular MOF phase forms over others is very limited, for example if kinetic factors associated with a particular MOF phase are prohibiting the synthesis, these databases will not (yet) have any information on this.

H1 Computational Discovery Studies.

When considering gas storage or separation applications, the availability of these databases allows us to analyse thousands to hundreds-of-thousands of materials such that general rules or guidelines for gas adsorption behaviour can be observed, providing insight for future experimental efforts. In general, these studies highlight important general physical characteristics, such as void fraction or surface area, that correlate to strong performance in a given application. These studies have provided invaluable insight into the remarkable potential nanoporous materials have, and some of their limitations. The following discussion has been split into several subsections that focus on a few examples. Table 1 gives a more complete overview, which includes the database of materials used, the topological diversity of these materials, and the gas conditions evaluated.

H2 Methane storage.

Methane storage capacity is by far the most studied sorption property in the field of computational high throughput screening of nanoporous materials[27–37,83]. There are several reasons for this, the first being that there is considerable interest in safely and efficiently storing methane for use as an alternative, clean-burning fuel in motor vehicles. To make storage in a porous material competitive with compressed natural gas, the US Department of Energy (DOE) has provided a target deliverable capacity of 315 volume

of methane per unit volume of material (vol %)[97] at adsorption pressures of 65 bar. This storage capacity is the energy density equivalent of storing methane in an empty tank at 200 bar, a pressure which is considerably expensive to support in a vehicle. This provides researchers with a clear goal to strive for Finally, it is worth noting that computing methane adsorption in nanoporous materials is one of the most accurate sorption simulations one can perform without resorting to more involved computational models, with some exceptions[98–100]. This is due to the isotropic and relatively non-polarizable nature of the methane particle, which can be easily modelled as either five rigid van der Waals particles located on each atom, or as a single unified van der Waals particle.

The main conclusions reached by these studies are similar, as many of them include simulating methane adsorption across a similar range of porous materials. For example, it was found that a region of high performing materials possessed largest cavity diameters of 10 – 12 Å[31,36], a region which was later expanded to 8.0 – 14.5 Å when a wider breadth of materials was considered[30]. Likewise, several studies show that materials with volumetric surface areas of around 2100 - 2300 $m^2 \cdot cm^{-3}$ and mild methane heats of adsorption (~11 $kJ \cdot mol^{-1}$) are found to have the highest deliverable capacities. It is remarkable to note that, in light of these exhaustive studies, it was realised one simply cannot obtain the DOE target of 315 vol % using nanoporous materials as the adsorbing media. This was proven recently[35], when the target could not be obtained even when artificially minimizing the density of framework atoms and maximizing methane adsorption. These artificial carbon frameworks are shown in Figure 4.

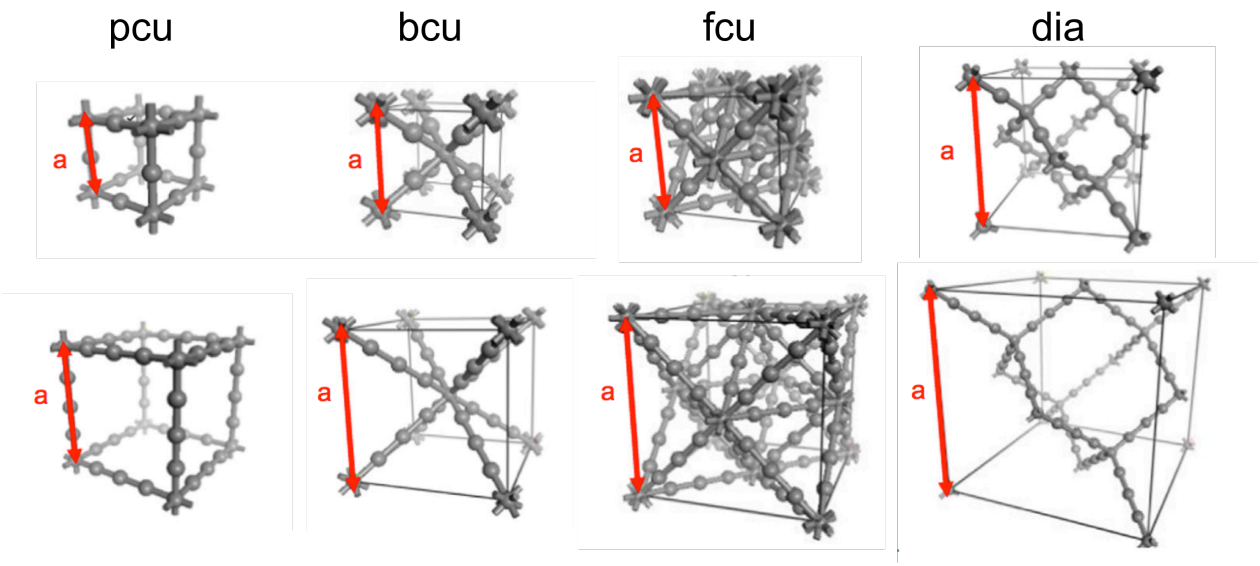

*Figure 4: Fictitious frameworks used to demonstrate the difficulty in obtaining the DOE target of 315 vol% deliverable capacity of methane. The highest these extremely high void space, carbon-based materials could reach was 80% of the target (the **dia** structure in the top right corner). Image adapted with permission from ref* [35].

The results from this body of work exemplify the reason for the creation of the nanoporous materials genome; An exhaustive search for the ideal material was performed across an enormous range of materials, the results of which found the maximum deliverable capacity of methane could reach 60% of the desired target. This motivated a thorough investigation on materials *beyond* what was considered synthetically feasible, and demonstrated that achieving the target deliverable capacity could only be obtained when extremely unphysical structures were used. This insight has provided a more realistic goal for potential commercial implementation, as there are still considerable advantages of having nano-porous materials store less methane in a tank at 65 bar, than a reinforced cylinder, using multi-stage compressors to reach the desired 200 bar pressure.

To enhance the expected deliverable capacity of these materials is always possible to adjust the thermodynamic conditions of gas cycling to enhance the amount of methane recovered from the adsorption process (for example increasing the temperature to burn off any residual methane)[37], albeit at higher costs than initially anticipated. Moreover, there is significant interest in flexible materials,

which possess steep steps in their methane adsorption isotherms when the material expands and contracts[101].

As a final point for discussion, nearly all of these screening studies for methane deliverable capacity make correlations between geometric properties of the materials and high performance. The properties looked at are typically simplified descriptors of a very complex manifold, such as a single numerical value representing the surface area, or the diameter of the largest cavity in a material (which assumes the cavity is a perfect sphere). The simplified nature of these descriptors limits their ability to capture important sorption characteristics within these materials, however it would seem that combining them, in the case of methane, was somewhat successful. In the following sections, we will witness that as chemistry begins to play more of a role in the behaviour of gas in nanopores, these geometric descriptors aren't nearly as predictive.

H2 Carbon Dioxide Sequestration.

Separating CO_2 from gas mixtures has become a major topic for materials screening studies[44,53,59,68,83,91]. It is believed that discovery-based design of new porous materials can provide major improvements to the costs of trapping CO_2 from mixed gas streams, and is linked to many applications from natural gas purification to post-combustion capture. While removal of CO_2 from natural gas reservoirs is a matured technology and extensively studied[102], the phenomena of separating CO_2 from waste production is relatively new, and is motivated by geopolitical efforts to reduce greenhouse gas production from major industrial contributors such as coal-fired power plants[103,104]. In this field, nanoporous materials could replace the incumbent technology, aqueous amines[105], by reducing the energetic costs to capture-regeneration cycles, while eliminating some of its undesirable side effects, such as the corrosive nature of the carbamate species formed from a chemical bond between amines and CO_2.

Unfortunately, unlike the specific thermodynamic conditions set for methane deliverable capacity, the targets for an ideal material in this category are much more open to interpretation, as they are typically expressed in terms of electricity costs for the consumer to a given process. However, two aspects of an ideal material are certain; the material needs to have a minimum selectivity for CO_2 over other gases in a mixture (called its selectivity), and the more CO_2 that can be recovered after an adsorption-desorption cycle (called its working capacity), the more optimal the material will be in an industrial setting. Without knowing the exact thermodynamic conditions or costs associated with capturing CO_2, several researchers have nevertheless attempted to provide metrics for estimating how well a nanoporous material could perform. For example, the sorption selectivity parameter[106] combines adsorption data from N_2 and CO_2 into a single value, while the parasitic energy[41] represents a measure of the loss of electricity production from a coal-fired power plant when equipped downstream with a CO_2 scrubber and compressor. The idea behind this metric is not only to reduce adsorption behaviour in materials to a single intuitive value, which allow us to rank materials, but also take into account that different materials have different optimal operating conditions.

It is fascinating to see the diverse structural designs suggested by computational screening studies in this field. For example it was demonstrated that ideal zeolitic materials can be found when optimal CO_2 adsorption pockets are present near large channels, a design which permits strong adsorption and fast diffusion[41]. Interestingly, no known zeolites possess this quality, which serves as a motivation for targeted zeolite synthesis efforts. Materials in the CoRE MOF database possessing alkaline or alkali metals are observed in 75% of the top performing materials (based on several metrics including CO_2:N_2 selectivity and CO_2 working capacity)[68]. While halogen functional groups were shown to have optimal performance for CO_2 capture in a screening of hypothetical MOFs[53]. An interesting result considering the unremarkable adsorption of CO_2 in some of the highest fluorinated MOFs reported in the literature[107,108].

These results have much to do with the breadth of materials being screened and the choice of the force field used (read – the interaction energy between the material and gas particle). The database of experimentally resolved MOFs (CoRE MOFs) possess a much wider range of chemistry and topologies than does the database of hypothetical MOFs, despite the difference in quantity (4764 CoRE MOFs vs 137,000 hMOFs). It is notable, however that many studies agree on the optimal energy of CO_2 adsorption; optimal materials are found when the isosteric heat of adsorption for CO_2 is between 28-35 $kJ \cdot mol^{-1}$ for CO_2:N_2 separations and 17-24 $kJ \cdot mol^{-1}$ for CO_2:CH_4 separations (a necessary, but not sufficient condition)[44,53,68].

In lieu of brute-force screening studies, several works have demonstrated that more advanced search techniques such as a genetic algorithm (GA) can be used to identify top performing materials[59,83,91]. In these algorithms, the focus is on refining the search for a good material by modifying the materials 'genes' so that it becomes superior in later generations. What these genes are, and how they are modified will ultimately determine the resulting MOF design guidelines of the search method. For example a GA was used to search for the optimal material for CO_2:H_2 separations[59]. Here, the genes of each hypothetical MOF were their building blocks: namely their functional groups, SBUs, and degree of interpenetration (a measure of how many times a MOF can interlink with itself). Each MOF's genes were mutated or mixed with other MOFs to form new MOFs (called children) until the final population showed very little improvement in their overall adsorption behaviour. By using a GA, they were able to identify a high performing material with ethoxy functional groups that was then successfully synthesised, NOTT-101-OEt[59], or Cu_2TPTC-OEt as it was called by the original discoverers[109].

It becomes apparent from these screening studies that geometric features such as material surface areas and void fractions are unable to capture the phenomena of adsorption for a relatively complex particle like CO_2. From a simulation perspective it is remarkable to have a seemingly subtle property such as CO_2's quadrupole moment become such a sweeping concern for high throughput simulations. What we

can learn from these studies is that when both chemistry and pore shape dominate sorption behaviour, we must look to different characteristics in order to aid experimental design. One approach is to identify regions of strong adsorption in high performing materials and extract the chemical features from those sites[49,110,111], however this may provide a too narrow view from a global search perspective.

H2 Other Gas Separations.

While methane and carbon dioxide have been the subject of many screening studies in the past decade, it is worth noting that nanoporous materials databases have been mined for other gas separations as well. For example several groups[49,54,112,113] have screened nanoporous materials databases for Xe:Kr separation, an important application for isolating isotopes from nuclear waste. These studies show that materials with pore sizes that *just* fit a Xe particle will be the most effective at separating Xe from Kr[54,113], which is the range between 4-8 Å. What is interesting from the graph of selectivity vs pore diameter shown in Figure 5a is how many structures *are not* selective with the optimal pore diameter, shown by the many points under the peak on the left-most part of the graph. What can be interpreted from this result is that, even for adsorption of single-atom particles such as Xe and Kr, robust design principles for these complex materials are difficult to identify with simple 1-dimensional pore descriptors. In lieu of providing design suggestions, one can scan a database of synthesizable materials for the top performers, and simply suggest them as targets for synthesis[59,65]. This is precisely what was done recently to identify and experimentally confirm the incredible Xe:Kr selectivity of SBMOF-1[113].

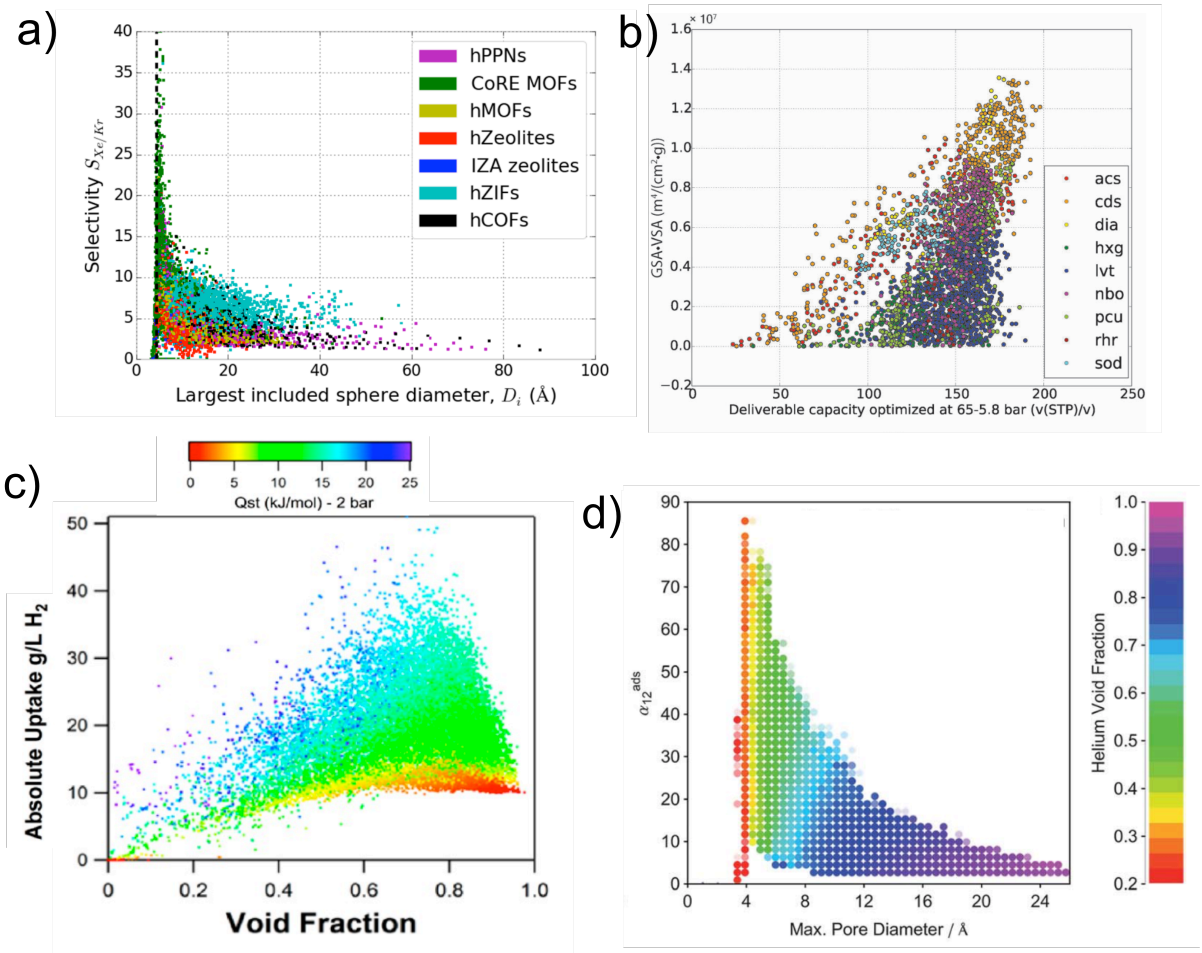

Figure 5: Examples of correlations between physical pore characteristics and performance. a) A plot of Xe:Kr selectivity vs pore diameter for a wide range of nanoporous materials at 1 bar and 298 K with a 20:80 molar Xe:Kr mixture. It is clear that the diameter must be around 4 Å to obtain high selectivity in these materials, however a large distribution of materials with little selectivity are observed at this diameter. b) Methane deliverable capacity is plotted against a measure of the surface area for a topologically diverse set of MOFs. The materials with the highest deliverable capacities could possess a complete range of surface area values. c) the adsorption of H_2 at 100 bar pressure and 248 K plotted as a function of material void fraction. While a clear peak is observed at ~0.75 void fraction, many materials perform poorly at this value. Each point is color-coded by its isosteric heat of adsorption at the desorption conditions of H_2 (2 bar and 243 K) d) the CO_2:N_2 selectivity vs pore diameter which has been mapped by void fraction in the third dimension. A small pore diameter of 4 Å and 0.2 void fraction appear as the optimal values for CO_2 selectivity, though a large number of materials perform poorly with these properties. a) was reproduced with permission from ref [49] b) was reproduced with permission from ref [83], c) was adapted with permission from ref [57], d) was reproduced with permission from ref [53].

In another case of clear DOE-assigned targets, H_2 storage has been the subject of screening studies[56,57,84,114] for the possibility of enhanced volumetric capacity of nanoporous materials. The DOE has set a target weight deliverable capacity of 7.5% mass of H_2 per mass of adsorbent and a volumetric deliverable density of 70 g·L^{-1} H_2[115]. It was shown that the adsorption of H_2 gas in porous frameworks was too weak to capture enough hydrogen at the adsorption pressure, thus a database of hypothetical

structures was constructed containing exposed Mg^{2+} metal ions[57], which provide strong adsorption sites for H_2. Upon screening the Mg^{2+} decorated materials, a density of 2.5 $mmol \cdot cm^{-3}$ Mg^{2+} ions in a framework was found to be ideal to provide a volumetric deliverable capacity of roughly 8 wt % and 30 $g \cdot L^{-1}$ H_2. While the volumetric requirement was not met, these materials show promise for delivering the desired amount of power to a motor vehicle per unit weight of the sorbent.

H1 Charge Generation

One of the major uncertainties in high throughput screening studies are the accuracy of the models describing the gas – framework interactions.[100,116–118] To capture the charge interactions between a gas particle and a materials' charge density, partial charges are typically assigned to each atom in the gas and framework. This effectively treats the complex electron density and nuclei of each atom as a single point charge positioned at its centre, which is then used to compute energy interactions with coulombs' law. The decision on how these point charges are assigned will play a major role in the accuracy of the results. For a handful of structures one can afford to perform individual DFT calculations on each, but high-throughput screening hundreds of thousands of structures is too computationally demanding. Therefore, it is an interesting computational challenge to develop more efficient alternative methods. For example, highly parameterised *ad hoc* methods such as the Charge Equilibration method (QEq)[119] or more recently the *Extended* Charge Equilibration Method (EQEq)[120,121] have been used to assign charges to framework atoms in simulation studies[53,55,59,122,123]. It was found that partial atomic charges, particularly for fluorine, were extremely overestimated by the QEq and EQEq methods.[124] The EQEq method was used to assign MOF atomic charges in some screening studies[53,55] which may have given rise to the noticeable influence of fluorine functionalised structures in the dissemination of the results.

In the QEq method the partial atomic charge generation can be extremely rapid, being on the order of seconds vs hours using *ab initio* calculations[125–127]. This is a vital aspect of a screening protocol when

simulating hundreds of thousands of materials. Using methods such as the Density Derived Electrostatic and Chemical (DDEC) charge method, which fits point charges based on *ab initio* atomic volumes and electrostatic potential[126,127], or the Repeating Electrostatic Potential Extracted Atomic charges (REPEAT) method[125], which fits atomic charges to the DFT electrostatic potential, can consume hours of compute time for *each material*. [125] Recognizing the discrepancy between rapid parameterised and slow *ab initio* based atomic charges, a new set of QEq parameters were developed called the MOF electrostatic-potential-optimized (MEPO) QEq parameters, such that the method could produce *ab initio* quality atomic charges for a select set of materials[124]. In that work, the authors warn that applying the MEPO-QEq method to systems that were not similar to the materials used in that study could result in spurious charge assignment. The nature of this method makes it questionable whether it can be applied directly to other materials[68].

In recognition of some of the limitations of parameterised methods, the group of David Sholl has introduced DFT-based DDEC charges for 2932 of the 4519 CoRE MOFs[128]. This was a substantial undertaking considering the cost of computing these charges, however it provides researchers with the tools necessary for screening these materials in a variety of applications where the guest-framework charge interactions are important. Moreover, it gives researchers a set of accurate charges on a wide variety of materials, ideal for developing new parameterised methods for fast charge generation using a robust training set.

H1 Data Mining Approaches.

When looking at these screening studies as a whole, one notices a trend in the general approach; A database is used or constructed with chemically meaningful constituents, followed by screening for a thermodynamic property, and finally structure-property relationships are examined from 2 or 3-

dimensional plots of performance as a function of material descriptors. The descriptors include, but are not limited to pore dimensionality, surface area, and void fraction, which can be calculated with several programs available for academic use[89,90,129,130]. Because of the dimensionality reduction of these descriptors from the complex 3-dimensional pores of a material to one-dimensional values, the relationships are usually not straight-forward.

One of the most important tests one can perform on these relations between material descriptor and performance is to determine how *predictive* they are, *i.e.* if we are given a structure, can we predict its performance, say in CO_2 separations, based on its surface area, pore size, or pore volume? There have been several recent studies dedicated to answering this question using models developed by machine learning[32,45,114,131–134], a field that is becoming extremely powerful in materials science[135,136]. It was shown that 1-dimensional geometric descriptors are able to successfully predict adsorption at high pressures[134,137] and low temperatures[114], using representative datasets of materials to train machine learning models. These models are however, rather poor at predicting performance at lower gas densities, in pressure regions of 0 – 1 bar[134]. This is likely due to the oversimplified nature of these descriptors, when a more detailed representation of the chemistry and pore shape is needed. Indeed, the development of complex materials descriptors that include both chemical and geometric features[134] appears to improve the overall success of machine learning models. Notably, however, the model in Ref 134 was poor at predicting adsorption performance on a certain materials containing chemical species such as F, Zr, and V[134]. This was due to the lack of representation of these materials in the dataset of MOFs used to train the SVM model, which raises an extremely important implication on materials databases moving forward; materials diversity. Due to the complex landscape of adsorbate-adsorbent interactions, it is currently unclear what the bounds are on materials diversity, and how many

descriptors are needed to quantify it. We feel that this is an issue that should be aggressively researched in the future of the materials genome project.

One promising avenue with respect to identifying diversity in these materials is to develop descriptors that capture the important shape and volume aspects of their pores. It was shown recently that using a topological data analysis technique to recognize and compare the pore shapes of each material, Lee *et al.* were able to not only identify materials with similar pores (that would be nearly impossible to do with the naked eye or conventional descriptors), but also capture trends in methane deliverable capacity in nanoporous materials by grouping similar materials together[138]. This tool presents an enormous opportunity for use in future machine-learning studies, as well as database curation.

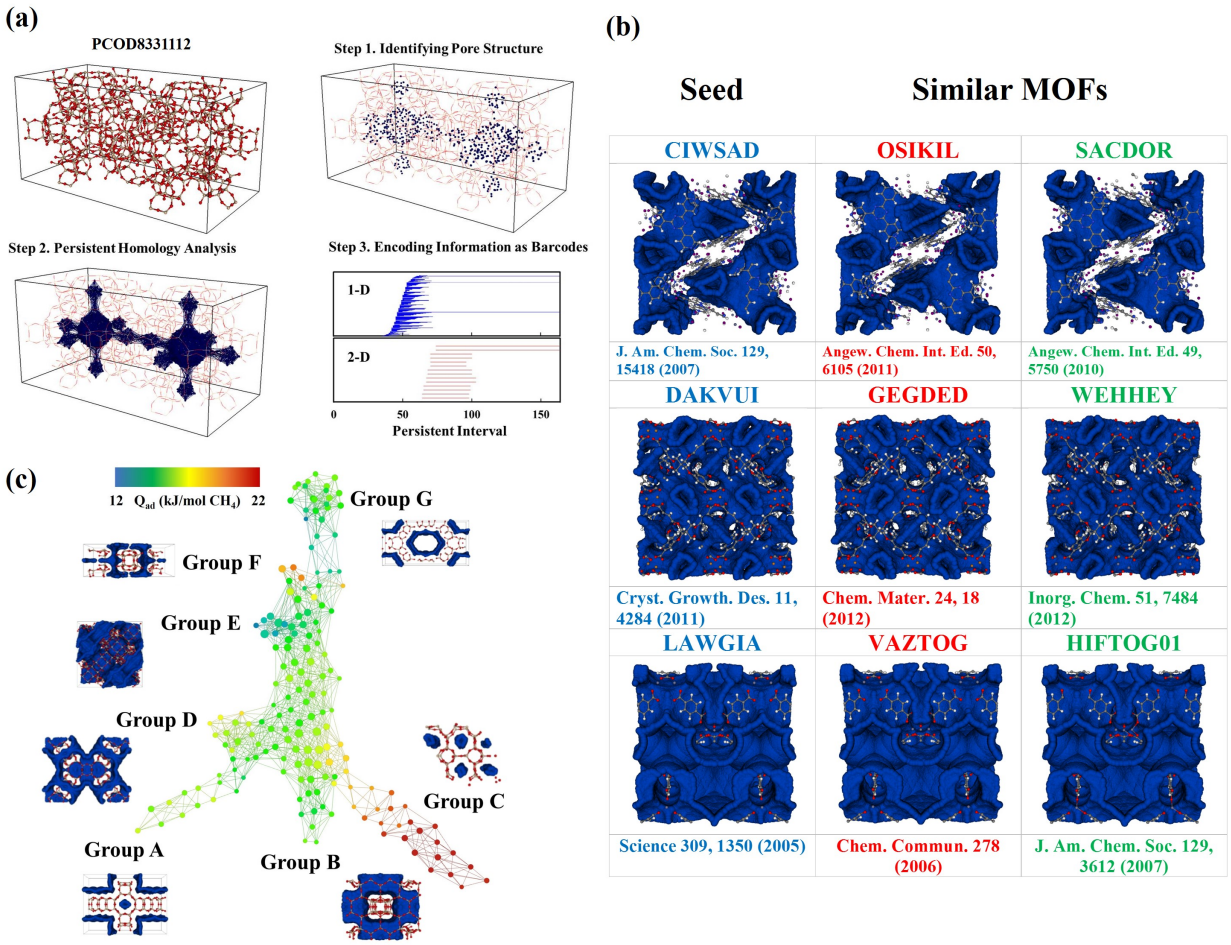

Figure 6: Development and usages of a descriptor using topological data analysis technique. a) Procedures to develop a descriptor based on pore shapes in nanoporous materials. The first step is preparation of a set of points to represent pore structure inside a nanoporous material using the open-source software Zeo++ [89,90]. Next, using the mathematical concept of persistent homology[139], the set of points are analyzed and the resulting information about pore structures are encoded as barcodes. The barcodes play a role as a descriptor for identifying and comparing materials. b) Materials from the CoRE-MOF database that have a similar pore geometry. Each row gives examples of materials that are very similar. The ones that are listed are those in which there are no cross references in the original articles of the corresponding similar structures c) Mapper plot generated by performing Topological Data Analysis on the subgroups of top-performing zeolites (top 1%) for methane storage application. Nodes in the network represent clusters of materials with similar pore shapes and edges connect nodes that contain structures in common. Nodes are colored by the average value of the heats of adsorption of the materials in a cluster (Red: high value, Blue: low value). It is shown examples from the seven topologically different classes of top-performing zeolites for methane storage. b) and c were reproduced with permission from ref [138].

H1 Outlook and Conclusions

In this review we have highlighted the current efforts made to both develop and utilize the nanoporous materials genome initiative. Databases of materials totaling over 3 million structures including MOFs,

COFs, Zeolites, ZIFs, and PPNs developed *in silico*, or extracted from the Cambridge Structural Database have been collected under this program to provide a chemically and topologically diverse range of materials. The overarching goal of this initiative is the broad concept of "materials discovery", which brings to mind many interpretations for a given application. We can narrow this to three main concepts; that of identifying candidates for synthesis, identifying transferrable properties for rational design, and identifying limits to performance. All of which have been tackled in a handful of gas storage applications. Some of the insight gained from this field of study has already been invaluable in the fields of CO_2 sequestration, methane and hydrogen storage.

As this field begins to reach maturity, several challenges still remain. Recent methods for enumerating hypothetical structures based on topological blueprints have provided rapid deployment of new topologically and chemically diverse materials. However, the approach of using SBUs as building blocks may be too simplistic, considering the coordination geometries observed in a majority of synthesised MOF materials appear to lack the typical geometric shapes of some widely studied organo-metallic SBUs[140]. Future efforts in this field will hopefully accommodate the possibility of the vast coordination environments exhibited by metallic ions in these structures, to further enhance the chemical diversity of the materials genome database.

An interesting question is whether these databases can be used to obtain a better understanding of the synthesis of MOFs. For example, if we take the case of ZIFs we can generate *in silico* for a given ZIF linker-metal combination any of the known zeolite or hypothetical zeolite structures. However, we cannot predict which of these millions of possible structures will form experimentally. While there are examples of focused studies on a handful of materials[141], if more data on synthesis conditions of these materials become available, big-data analysis tools may give us some insights in possible correlations between synthesis conditions and the structure that will form.

With the novel tools developed to enumerate these materials, we now have a path to arguably *too* many materials. The focus must now switch from collection to curation; how to develop a diverse, and importantly, evenly distributed database of materials for future screening efforts. Ensuring quality over quantity is not an easy task, when diversity bounds and differences between materials are still an unknown quantity. Future efforts towards identifying material features that, independent of a given application, can effectively categorize and differentiate these structures must take top priority.

The studies discussed in this review have focused on small-molecule gas adsorption. In these cases, it is computationally efficient to model gas adsorption in these materials while assuming the atoms in the framework do not move (rigid approximation). One can argue that these studies are the low hanging fruit; the databases were still relatively small and the calculation sufficiently efficient that brute-force simulation could be used. The next step will give interesting computational challenges where brute force simulations won't be sufficient. For example, we would like to query these databases for more complex behaviour, the rigid approximation must be relaxed. It has been recognised that modeling flexibility is of paramount importance when screening materials[61], particularly when their pore sizes are similar to the diameter of a gas particle[142–144], or when modeling the breathing phenomena exhibited by some of these materials[145–149]. It is known, for example, that the mechanical stability of MOFs are generally worse than that of zeolites and dense hybrid materials, affecting their commercial-scale implementation[150]. Computational studies on the mechanical stability of these materials are still quite focused and typically require quantum chemical calculations to provide accurate results[86]. Moreover many of these materials are sensitive to water, and while one can compute a measure of hydrophobicity in these materials to give an idea of their permanence in water[66], investigating chemical pathways to water degredation can be extremely computationally expensive for a single material[151–154]. While accurate calculations of water stability are currently too expensive for these materials databases, creative development of chemical

descriptors that correlate well with observed water degradation trends would fill an important knowledge gap.

These properties introduce a broad range of challenges, the primary of which is to accurately model time-dependent thermodynamic behaviour with computationally cheap classical force fields developed specifically for nanoporous materials. We are beginning to witness efforts towards developing generalized force fields for these applications[155–159].

More ambitious applications for these databases are to characterize their performance in other, emerging applications such as catalysis or sensing. For these applications one needs an understanding on how porous materials respond to external stimuli. These types of calculations require state of art quantum calculations, which are at present too time consuming for any large-scale screening. In addition, the number of new materials will most likely grow faster than our capacity for brute-force screening. It will be interesting to see how this will be solved with the development of novel sampling and low-cost simulation techniques, or the application of big-data methods.

In this review we have been critical on many aspects, but we would like to conclude with the observation that the collection of all databases of nanoporous materials represents the most complete and sophisticated knowledge base of possible variations of known materials. Also, the molecular simulation techniques that have been used to make predictions have been validated such that for the large majority of materials very reasonable predictions can be made. If one would like to know what the maximum performance of these materials for a give application, a screening of say, 100,000 materials will give us the best estimate what the current state of the art can bring us. It will be interesting to see whether a large-scale computational screening will become a commonly-accepted first step to evaluate the potential of a class of materials.

In the near future, complete databases will be made publicly available to encourage researchers to use the data for their own agendas. To date, only a selection of external researchers have used this extensive and extremely useful resource, but it is these authors' hope that there will be further participation going forward. These databases will soon be publically available on the Materials Cloud (http://materialscloud.org/archive/) to foster collaboration and the distribution of knowledge. It is anticipated that experimental groups will be able to upload their new structures and compare with big-data predictions of similar structures, while suggestions can be given for promising applications of these materials.

Acknowledgement

This work is supported by Swiss National Science Foundation (SNSF) through the National Center of Competence in Research (NCCR) ' Materials' Revolution: Computational Design and Discovery of Novel Materials (MARVEL)' and the Korean-Swiss Science and Technology Programme (KSSTP, grant number 162130) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 666983, MaGic).

H1 References

1. Flanigen, E. M., Broach, R. W. & Wilson, S. T. in *Zeolites in Industrial Separation and Catalysis* 1–26 (Wiley-VCH Verlag GmbH & Co. KGaA, 2010). doi:10.1002/9783527629565.ch1

2. Zimmermann, N. E. R. & Haranczyk, M. History and Utility of Zeolite Framework-Type Discovery from a Data-Science Perspective. *Cryst. Growth Des.* **16,** 3043–3048 (2016).

3. International Zeolite Association. at <http://www.iza-online.org>

4. Hoskin, B. F. & Robson, R. Design and construction of a new class of scaffolding-like materials comprising infinite polymeric frameworks of 3D-linked molecular rods. A reappraisal of. *J. Am. Chem. Soc.* **112,** 1546–1554 (1990).

5. Chui, S. S.-Y., Lo, S. M.-F., Charmant, J. P. H., Orpen, A. G. & Williams, I. D. A Chemically Functionalizable Nanoporous Material [Cu3(TMA)2(H2O)3]n. *Science (80-.).* **283,** 1148–1150 (1999).

6. Li, H., Eddaoudi, M., O'Keeffe, M. & Yaghi, O. M. Design and synthesis of an exceptionally stable and highly porous metal-organic framework. *Nature* **402,** 276–279 (1999).

7. Yaghi, O. M. & Li, H. Hydrothermal Synthesis of a Metal-Organic Framework Containing Large Rectangular Channels. *J. Am. Chem. Soc.* **117,** 10401–10402 (1995).

8. Batten, S. R. *et al.* Terminology of metal–organic frameworks and coordination polymers (IUPAC Recommendations 2013). *Pure Appl. Chem.* **85,** 1715–1724 (2013).

9. Côté, A. P. *et al.* Porous, crystalline, covalent organic frameworks. *Science* **310,** 1166–1170 (2005).

10. El-Kaderi, H. M. *et al.* Designed Synthesis of 3D Covalent Organic Frameworks. *Science (80-.).* **316,** 268–272 (2007).

11. Banerjee, R. *et al.* High-throughput synthesis of zeolitic imidazolate frameworks and application to CO2 capture. *Science* **319,** 939–943 (2008).

12. Lu, W. *et al.* Porous Polymer Networks: Synthesis, Porosity, and Applications in Gas Storage/Separation. *Chem. Mater.* **22,** 5964–5972 (2010).

13. Yaghi, O. M. *et al.* Reticular synthesis and the design of new materials. *Nature* **423,** 705–714 (2003).

14. Ockwig, N. W., Delgado-Friedrichs, O., O'Keeffe, M. & Yaghi, O. M. Reticular Chemistry: Occurrence and Taxonomy of Nets and Grammar for the Design of Frameworks. *Acc. Chem. Res.* **38,** 176–182 (2005).

15. Delgado-Friedrichs, O., O'Keeffe, M. & Yaghi, O. M. Taxonomy of periodic nets and the design of materials. *Phys. Chem. Chem. Phys.* **9,** 1035–43 (2007).

16. Bonneau, C., Delgado-Friedrichs, O., O'Keeffe, M. & Yaghi, O. M. Three-periodic nets and tilings: minimal nets. *Acta Crystallogr. A.* **60,** 517–20 (2004).

17. Delgado-Friedrichs, O. & O'Keeffe, M. Crystal nets as graphs: Terminology and definitions. *J. Solid State Chem.* **178,** 2480–2485 (2005).

18. Li, M., Li, D., O'Keeffe, M. & Yaghi, O. M. Topological Analysis of Metal–Organic Frameworks with Polytopic Linkers and/or Multiple Building Units and the Minimal Transitivity Principle. *Chem. Rev.* **114,** 1343–1370 (2014).

19. O'Keeffe, M. & Yaghi, O. M. Deconstructing the Crystal Structures of Metal–Organic Frameworks and Related Materials into Their Underlying Nets. *Chem. Rev.* **112,** 675–702 (2012).

20. Schoedel, A., Li, M., Li, D., O'Keeffe, M. & Yaghi, O. M. Structures of Metal–Organic Frameworks with Rod Secondary Building Units. *Chem. Rev.* **116,** 12466–12535 (2016).

21. Blatov, V. a., Carlucci, L., Ciani, G. & Proserpio, D. M. Interpenetrating metal organic and inorganic 3D networks: a computer-aided systematic investigation. Part I. Analysis of the Cambridge structural database. *CrystEngComm* **6,** 378 (2004).

22. Alexandrov, E. V., Blatov, V. A., Kochetkov, A. V. & Proserpio, D. M. Underlying nets in three-periodic coordination polymers: topology, taxonomy and prediction from a computer-aided analysis of the Cambridge Structural Database. *CrystEngComm* **13,** 3947 (2011).

23. Farha, O. K. *et al.* De novo synthesis of a metal-organic framework material featuring ultrahigh surface area and gas storage capacities. *Nat. Chem.* **2,** 944–8 (2010).

24. Vaidhyanathan, R. *et al.* Direct observation and quantification of CO_2 binding within an amine-functionalized nanoporous solid. *Science* **330,** 650–3 (2010).

25. McDonald, T. M. *et al.* Cooperative insertion of CO2 in diamine-appended metal-organic frameworks. *Nature* **519,** 303–308 (2015).

26. Lyne, P. D. Structure-based virtual screening: an overview. *Drug Discov. Today* **7,** 1047–1055 (2002).

27. Colón, Y. J. & Snurr, R. Q. High-throughput computational screening of metal–organic frameworks. *Chem. Soc. Rev.* **43,** 5735–5749 (2014).

28. Martin, R. L. *et al.* In Silico Design of Three-Dimensional Porous Covalent Organic Frameworks via Known Synthesis Routes and Commercially Available Species. *J. Phys. Chem. C* **118,** 23790–23802 (2014).

29. Simon, C. M. *et al.* Optimizing nanoporous materials for gas storage. *Phys. Chem. Chem. Phys.* **16,** 5499–5513 (2014).

30. Simon, C. M. *et al.* The materials genome in action: identifying the performance limits for

methane storage. *Energy Environ. Sci.* **8,** 1190–1199 (2015).

31. Martin, R. L., Simon, C. M., Smit, B. & Haranczyk, M. In silico Design of Porous Polymer Networks: High-Throughput Screening for Methane Storage Materials. *J. Am. Chem. Soc.* **136,** 5006–22 (2014).

32. Ohno, H. & Mukae, Y. Machine Learning Approach for Prediction and Search: Application to Methane Storage in a Metal–Organic Framework. *J. Phys. Chem. C* **120,** 23963–23968 (2016).

33. Martin, R. L., Lin, L.-C., Jariwala, K., Smit, B. & Haranczyk, M. Mail-Order Metal–Organic Frameworks (MOFs): Designing Isoreticular MOF-5 Analogues Comprising Commercially Available Organic Molecules. *J. Phys. Chem. C* **117,** 12159–12167 (2013).

34. Kim, J. *et al.* New materials for methane capture from dilute and medium-concentration sources. *Nat. Commun.* **4,** 1694 (2013).

35. Gómez-Gualdrón, D. A. *et al.* Impact of the strength and spatial distribution of adsorption sites on methane deliverable capacity in nanoporous materials. *Chem. Eng. Sci.* **159,** 18–30 (2017).

36. Gómez-Gualdrón, D. A., Wilmer, C. E., Farha, O. K., Hupp, J. T. & Snurr, R. Q. Exploring the Limits of Methane Storage and Delivery in Nanoporous Materials. *J. Phys. Chem. C* **118,** 6941–6951 (2014).

37. Fu, J., Tian, Y. & Wu, J. Seeking metal–organic frameworks for methane storage in natural gas vehicles. *Adsorption* **21,** 499–507 (2015).

38. Krishna, R. & van Baten, J. M. In silico screening of metal–organic frameworks in separation applications. *Phys. Chem. Chem. Phys.* **13,** 10593 (2011).

39. Chung, Y. G. *et al.* Computation-Ready, Experimental Metal–Organic Frameworks: A Tool To Enable High-Throughput Screening of Nanoporous Crystals. *Chem. Mater.* **26,** 6185–6192 (2014).

40. Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr. Sect. B Struct. Sci.* **58,** 380–388 (2002).

41. Lin, L.-C. *et al.* In silico screening of carbon-capture materials. *Nat. Mater.* **11,** 633–41 (2012).

42. Kim, J., Lin, L., Swisher, J. A., Haranczyk, M. & Smit, B. Predicting Large CO 2 Adsorption in Aluminosilicate Zeolites for Postcombustion Carbon Dioxide Capture. *J. Am. Chem. Soc.* **134,** 18940–18943 (2012).

43. Kim, J., Abouelnasr, M., Lin, L.-C. & Smit, B. Large-Scale Screening of Zeolite Structures for CO 2 Membrane Separations. *J. Am. Chem. Soc.* **135,** 7545–7552 (2013).

44. Braun, E. *et al.* High-throughput computational screening of nanoporous adsorbents for CO 2 capture from natural gas. *Mol. Syst. Des. Eng.* **1,** 175–188 (2016).

45. Thornton, A. W., Winkler, D. A., Liu, M. S., Haranczyk, M. & Kennedy, D. F. Towards Computational Design of Zeolite Catalysts for CO2 Reduction. *RSC Adv.* **5,** 44361–44370 (2015).

46. Jeong, W. & Kim, J. Understanding the Mechanisms of CO 2 Adsorption Enhancement in Pure Silica Zeolites under Humid Conditions. *J. Phys. Chem. C* **120,** 23500–23510 (2016).

47. Kim, J. *et al.* Large-Scale Computational Screening of Zeolites for Ethane/Ethene Separation. *Langmuir* **28,** 11914–11919 (2012).

48. Bai, P. *et al.* Discovery of optimal zeolites for challenging separations and chemical transformations using predictive materials modeling. *Nat. Commun.* **6,** 5912 (2015).

49. Simon, C. M., Mercado, R., Schnell, S. K., Smit, B. & Haranczyk, M. What Are the Best Materials To Separate a Xenon/Krypton Mixture? *Chem. Mater.* **27,** 4459–4475 (2015).

50. Poloni, R. & Kim, J. Predicting low-k zeolite materials. *J. Mater. Chem. C* **2,** 2298 (2014).

51. Wilmer, C. E. *et al.* Large-scale screening of hypothetical metal-organic frameworks. *Nat. Chem.* **4,** 83–9 (2012).

52. Sikora, B. J., Winnegar, R., Proserpio, D. M. & Snurr, R. Q. Textural properties of a large collection of computationally constructed MOFs and zeolites. *Microporous Mesoporous Mater.* **186,** 207–213 (2014).

53. Wilmer, C. E., Farha, O. K., Bae, Y.-S., Hupp, J. T. & Snurr, R. Q. Structure–property relationships of porous materials for carbon dioxide separation and capture. *Energy Environ. Sci.* **5,** 9849 (2012).

54. Sikora, B. J., Wilmer, C. E., Greenfield, M. L. & Snurr, R. Q. Thermodynamic analysis of Xe/Kr selectivity in over 137 000 hypothetical metal–organic frameworks. *Chem. Sci.* **3,** 2217 (2012).

55. Moghadam, P. Z., Fairen-Jimenez, D. & Snurr, R. Q. Efficient identification of hydrophobic MOFs: application in the capture of toxic industrial chemicals. *J. Mater. Chem. A* **4,** 529–536 (2016).

56. Bobbitt, N. S., Chen, J. & Snurr, R. Q. High-Throughput Screening of Metal–Organic Frameworks for Hydrogen Storage at Cryogenic Temperature. *J. Phys. Chem. C* **120,** 27328–27341 (2016).

57. Colón, Y. J., Fairen-Jimenez, D., Wilmer, C. E. & Snurr, R. Q. High-Throughput Screening of Porous Crystalline Materials for Hydrogen Storage Capacity near Room Temperature. *J. Phys. Chem. C* **118,** 5383–5389 (2014).

58. Gomez, D. A., Toda, J. & Sastre, G. Screening of hypothetical metal–organic frameworks for H 2 storage. *Phys. Chem. Chem. Phys.* **16,** 19001 (2014).

59. Chung, Y. G. *et al.* In silico discovery of metal-organic frameworks for precombustion CO2 capture using a genetic algorithm. *Sci. Adv.* **2,** e1600909–e1600909 (2016).

60. Qiao, Z., Peng, C., Zhou, J. & Jiang, J. High-throughput computational screening of 137953 metal–organic frameworks for membrane separation of a CO 2 /N 2 /CH 4 mixture. *J. Mater. Chem. A* **4,**

15904–15912 (2016).

61. Sarkisov, L., Martin, R. L., Haranczyk, M. & Smit, B. On the Flexibility of Metal–Organic Frameworks. *J. Am. Chem. Soc.* **136,** 2228–2231 (2014).

62. First, E. L., Gounaris, C. E. & Floudas, C. A. Predictive Framework for Shape-Selective Separations in Three-Dimensional Zeolites and Metal–Organic Frameworks. *Langmuir* **29,** 5599–5608 (2013).

63. Xiang, Z. *et al.* Systematic Tuning and Multifunctionalization of Covalent Organic Polymers for Enhanced Carbon Capture. *J. Am. Chem. Soc.* **137,** 13301–13307 (2015).

64. Witman, M. *et al.* In silico design and screening of hypothetical MOF-74 analogs and their experimental synthesis. *Chem. Sci.* **7,** 6263–6272 (2016).

65. Gee, J. A. *et al.* Computational Identification and Experimental Evaluation of Metal–Organic Frameworks for Xylene Enrichment. *J. Phys. Chem. C* **120,** 12075–12082 (2016).

66. Li, S., Chung, Y. G. & Snurr, R. Q. High-Throughput Screening of Metal–Organic Frameworks for CO 2 Capture in the Presence of Water. *Langmuir* **32,** 10368–10376 (2016).

67. Yeo, B. C., Kim, D., Kim, H. & Han, S. S. High-Throughput Screening to Investigate the Relationship between the Selectivity and Working Capacity of Porous Materials for Propylene/Propane Adsorptive Separation. *J. Phys. Chem. C* acs.jpcc.6b08177 (2016). doi:10.1021/acs.jpcc.6b08177

68. Qiao, Z., Zhang, K. & Jiang, J. In silico screening of 4764 computation-ready, experimental metal–organic frameworks for CO 2 separation. *J. Mater. Chem. A* **4,** 2105–2114 (2016).

69. Van Heest, T., Teich-McGoldrick, S. L., Greathouse, J. a., Allendorf, M. D. & Sholl, D. S. Identification of Metal–Organic Framework Materials for Adsorption Separation of Rare Gases: Applicability of Ideal Adsorbed Solution Theory (IAST) and Effects of Inaccessible Framework Regions. *J. Phys. Chem. C* **116,** 13183–13195 (2012).

70. O'Keeffe, M., Peskov, M. A., Ramsden, S. J. & Yaghi, O. M. The Reticular Chemistry Structure Resource (RCSR) database of, and symbols for, crystal nets. *Acc. Chem. Res.* **41,** 1782–9 (2008).

71. Pophale, R., Cheeseman, P. a & Deem, M. W. A database of new zeolite-like materials. *Phys. Chem. Chem. Phys.* **13,** 12407–12412 (2011).

72. Deem, M. W., Pophale, R., Cheeseman, P. A. & Earl, D. J. Computational Discovery of New Zeolite-Like Materials. *J. Phys. Chem. C* **113,** 21353–21360 (2009).

73. Mellot-Draznieks, C., Dutour, J. & Férey, G. Hybrid Organic-Inorganic Frameworks: Routes for Computational Design and Structure Prediction. *Angew. Chemie Int. Ed.* **43,** 6290–6296 (2004).

74. Mellot Draznieks, C., Newsam, J. M., Gorman, A. M., Freeman, C. M. & Férey, G. De Novo Prediction of Inorganic Structures Developed through Automated Assembly of Secondary Building Units (AASBU Method). *Angew. Chemie Int. Ed.* **39,** 2270–2275 (2000).

75. Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. Optimization by Simulated Annealing. *Science (80-.).* **220,** 671–680 (1983).

76. Falcioni, M. & Deem, M. W. A biased Monte Carlo scheme for zeolite structure solution. *J. Chem. Phys.* **110,** 1754–1766 (1999).

77. Deem, M. W. & Newsam, J. M. Determination of 4-connected framework crystal structures by simulated annealing. *Nature* **342,** 260–262 (1989).

78. Mellot-Draznieks, C. *et al.* Computational design and prediction of interesting not-yet-synthesized structures of inorganic materials by using building unit concepts. *Chemistry - A European Journal* **8,** 4102–4113 (2002).

79. Chen, B., Eddaoudi, M., Hyde, S. T., O'Keeffe, M. & Yaghi, O. M. Interwoven metal-organic framework on a periodic minimal surface with extra-large pores. *Science* **291,** 1021–3 (2001).

80. Skiena, S. S. *The Algorithm Design Manual*. *Algorithm Design Manual (2nd Edition)* (2009). doi:10.1007/978-1-84800-070-4

81. Martin, R. L. & Haranczyk, M. Exploring frontiers of high surface area metal--organic frameworks. *Chem. Sci.* **4,** 1781–1785 (2013).

82. Martin, R. L. & Haranczyk, M. Optimization-Based Design of Metal–Organic Framework Materials. *J. Chem. Theory Comput.* **9,** 2816–2825 (2013).

83. Bao, Y., Martin, R. L., Haranczyk, M. & Deem, M. W. In silico prediction of MOFs with high deliverable capacity or internal surface area. *Phys. Chem. Chem. Phys.* **17,** 11962–11973 (2015).

84. Gómez-Gualdrón, D. A. *et al.* Evaluating topologically diverse metal–organic frameworks for cryo-adsorbed hydrogen storage. *Energy Environ. Sci.* **9,** 3279–3289 (2016).

85. Wang, T. C. *et al.* Ultrahigh Surface Area Zirconium MOFs and Insights into the Applicability of the BET Theory. *J. Am. Chem. Soc.* 150309080817005 (2015). doi:10.1021/ja512973b

86. Coudert, F.-X. & Fuchs, A. H. Computational characterization and prediction of metal–organic framework properties. *Coord. Chem. Rev.* **307,** 211–236 (2016).

87. Colón, Y. J. & Snurr, R. Q. High-throughput computational screening of metal-organic frameworks. *Chem. Soc. Rev.* 5735–5749 (2014). doi:10.1039/c4cs00070f

88. Martin, R. L. & Haranczyk, M. Construction and Characterization of Structure Models of Crystalline Porous Polymers. *Cryst. Growth Des.* **14,** 2431–2440 (2014).

89. Willems, T. F., Rycroft, C. H., Kazi, M., Meza, J. C. & Haranczyk, M. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.* **149,** 134–141 (2012).

90. Pinheiro, M. *et al.* Characterization and comparison of pore landscapes in crystalline porous materials. *J. Mol. Graph. Model.* **44,** 208–219 (2013).

91. Bao, Y. *et al.* In Silico Discovery of High Deliverable Capacity Metal–Organic Frameworks. *J. Phys. Chem. C* **119,** 186–195 (2015).

92. Addicoat, M. a, Coupry, D. E. & Heine, T. AuToGraFS: Automatic Topological Generator for Framework Structures. *J. Phys. Chem. A* **118,** 9607–14 (2014).

93. Gale, J. GULP: A computer program for the symmetry-adapted simulation of solids. *J. Chem. Soc., Faraday Trans.* **93,** 629–637 (1997).

94. Coupry, D. E., Addicoat, M. A. & Heine, T. Extension of the Universal Force Field for Metal–Organic Frameworks. *J. Chem. Theory Comput.* **12,** 5215–5225 (2016).

95. Addicoat, M. a., Vankova, N., Akter, I. F. & Heine, T. Extension of the Universal Force Field to Metal–Organic Frameworks. *J. Chem. Theory Comput.* **10,** 880–891 (2014).

96. Boyd, P. G. & Woo, T. K. A generalized method for constructing hypothetical nanoporous materials of any net topology from graph theory. *CrystEngComm* **18,** 3777–3792 (2016).

97. U.S. Department of Energy. *ARPA-E Methane Opportunities for Vehicular Energy (MOVE) (DE-FOA-000672)*. (2012). at <http://arpa-e-foa.energy.gov>

98. Hulvey, Z. *et al.* Critical Factors Driving the High Volumetric Uptake of Methane in Cu 3 (btc) 2. *J. Am. Chem. Soc.* **137,** 10816–10825 (2015).

99. Kim, J., Lin, L.-C., Lee, K., Neaton, J. B. & Smit, B. Efficient Determination of Accurate Force Fields for Porous Materials Using ab Initio Total Energy Calculations. *J. Phys. Chem. C* **118,** 2693–2701 (2014).

100. Becker, T. M., Heinen, J., Dubbeldam, D., Lin, L.-C. & Vlugt, T. J. H. Polarizable Force Fields for CO 2 and CH 4 Adsorption in M-MOF-74. *J. Phys. Chem. C* acs.jpcc.6b12052 (2017). doi:10.1021/acs.jpcc.6b12052

101. Mason, J. a. *et al.* Methane storage in flexible metal–organic frameworks with intrinsic thermal management. *Nature* **527,** 357–361 (2015).

102. Rufford, T. E. *et al.* The removal of CO2 and N2 from natural gas: A review of conventional and emerging process technologies. *J. Pet. Sci. Eng.* **94–95,** 123–154 (2012).

103. Lee, Z. H., Lee, K. T., Bhatia, S. & Mohamed, A. R. Post-combustion carbon dioxide capture: Evolution towards utilization of nanomaterials. *Renew. Sustain. Energy Rev.* **16,** 2599–2609 (2012).

104. Statistics, I. *CO2 emissions from fuel combustion-highlights*. *IEA, Paris http://www. iea. org/co2highlights/co2highlights.pdf* (2012). at <http://www.iea.org/co2highlights/co2highlights.pdf>

105. Abu-Zahra, M. R. M., Schneiders, L. H. J., Niederer, J. P. M., Feron, P. H. M. & Versteeg, G. F. CO2 capture from power plants. *Int. J. Greenh. Gas Control* **1,** 37–46 (2007).

106. Bae, Y.-S. & Snurr, R. Q. Development and Evaluation of Porous Materials for Carbon Dioxide Separation and Capture. *Angew. Chemie Int. Ed.* **50,** 11586–11596 (2011).

107. Chen, T.-H. *et al.* Mesoporous Fluorinated Metal-Organic Frameworks with Exceptional Adsorption of Fluorocarbons and CFCs. *Angew. Chemie Int. Ed.* **54,** 13902–13906 (2015).

108. Pachfule, P., Chen, Y., Sahoo, S. C., Jiang, J. & Banerjee, R. Structural Isomerism and Effect of Fluorination on Gas Adsorption in Copper-Tetrazolate Based Metal Organic Frameworks. *Chem. Mater.* **23,** 2908–2916 (2011).

109. Makal, T. A., Wang, X. & Zhou, H.-C. Tuning the Moisture and Thermal Stability of Metal–Organic Frameworks through Incorporation of Pendant Hydrophobic Groups. *Cryst. Growth Des.* **13,** 4760–4768 (2013).

110. Li, Z., Xiao, G., Yang, Q., Xiao, Y. & Zhong, C. Computational exploration of metal–organic frameworks for CO2/CH4 separation via temperature swing adsorption. *Chem. Eng. Sci.* **120,** 59–66 (2014).

111. Tong, M., Yang, Q., Xiao, Y. & Zhong, C. Revealing the structure–property relationship of covalent organic frameworks for CO2 capture from postcombustion gas: a multi-scale computational study. *Phys. Chem. Chem. Phys.* **16,** 15189 (2014).

112. Banerjee, D. *et al.* Metal–organic framework with optimally selective xenon adsorption and separation. *Nat. Commun.* **7,** 11831 (2016).

113. Sumer, Z. & Keskin, S. Molecular Simulations of MOF Adsorbents and Membranes for Noble Gas Separations. *Chem. Eng. Sci.* (2017). doi:10.1016/j.ces.2017.02.010

114. Thornton, A. W. *et al.* Materials Genome in Action: Identifying the Performance Limits of Physical Hydrogen Storage. *Chem. Mater.* **29,** 2844–2854 (2017).

115. U.S. Department of Energy. *DOE Targets for Onboard Hydrogen Storage Systems for Light-Duty Vehicles*. (2012). at <https://energy.gov/sites/prod/files/2015/05/f22/fcto_targets_onboard_hydro_storage_explanation.pdf>

116. McDaniel, J. G., Li, S., Tylianakis, E., Snurr, R. Q. & Schmidt, J. R. Evaluation of Force Field Performance for High-Throughput Screening of Gas Uptake in Metal–Organic Frameworks. *J. Phys. Chem. C* **119,** 3143–3152 (2015).

117. McDaniel, J. G. & Schmidt, J. R. Robust, Transferable, and Physically Motivated Force Fields for Gas Adsorption in Functionalized Zeolitic Imidazolate Frameworks. *J. Phys. Chem. C* **116,** 14031–14039 (2012).

118. Mercado, R. *et al.* Force Field Development from Periodic Density Functional Theory Calculations for Gas Separation Applications Using Metal–Organic Frameworks. *J. Phys. Chem. C* **120,** 12590–12604 (2016).

119. Rappe, A. K. & Goddard, W. A. Charge equilibration for molecular dynamics simulations. *J. Phys. Chem.* **95,** 3358–3363 (1991).

120. Wilmer, C. E. & Snurr, R. Q. Towards rapid computational screening of metal-organic frameworks for carbon dioxide capture: Calculation of framework charges via charge equilibration. *Chem. Eng. J.* **171,** 775–781 (2011).

121. Wilmer, C. E., Kim, K. C. & Snurr, R. Q. An Extended Charge Equilibration Method. *J. Phys. Chem. Lett.* **3,** 2506–2511 (2012).

122. Haldoupis, E., Nair, S. & Sholl, D. S. Finding MOFs for highly selective CO2/N2 adsorption using materials screening based on efficient assignment of atomic point charges. *J. Am. Chem. Soc.* **134,** 4313–23 (2012).

123. Wells, B. A., De Bruin-Dickason, C. & Chaffee, A. L. Charge Equilibration Based on Atomic Ionization in Metal–Organic Frameworks. *J. Phys. Chem. C* **119,** 456–466 (2015).

124. Kadantsev, E. S., Boyd, P. G., Daff, T. D. & Woo, T. K. Fast and Accurate Electrostatics in Metal Organic Frameworks with a Robust Charge Equilibration Parameterization for High-Throughput Virtual Screening of Gas Adsorption. *J. Phys. Chem. Lett.* **4,** 3056–3061 (2013).

125. Campañá, C., Mussard, B. & Woo, T. K. Electrostatic Potential Derived Atomic Charges for Periodic

Systems Using a Modified Error Functional. *J. Chem. Theory Comput.* **5,** 2866–2878 (2009).

126. Manz, T. A. & Sholl, D. S. Chemically Meaningful Atomic Charges That Reproduce the Electrostatic Potential in Periodic and Nonperiodic Materials. *J. Chem. Theory Comput.* **6,** 2455–2468 (2010).

127. Manz, T. A. & Sholl, D. S. Improved atoms-in-molecule charge partitioning functional for simultaneously reproducing the electrostatic potential and chemical states in periodic and nonperiodic materials. *J. Chem. Theory Comput.* **8,** 120718154902008 (2012).

128. Nazarian, D., Camp, J. S. & Sholl, D. S. A Comprehensive Set of High-Quality Point Charges for Simulations of Metal–Organic Frameworks. *Chem. Mater.* **28,** 785–793 (2016).

129. First, E. L., Gounaris, C. E., Wei, J. & Floudas, C. a. Computational characterization of zeolite porous networks: an automated approach. *Phys. Chem. Chem. Phys.* **13,** 17339–58 (2011).

130. Sarkisov, L. & Harrison, A. Computational structure characterisation tools in application to ordered and disordered porous materials. *Mol. Simul.* **37,** 1248–1257 (2011).

131. Aghaji, M. Z., Fernandez, M., Boyd, P. G., Daff, T. D. & Woo, T. K. Quantitative Structure-Property Relationship Models for Recognizing Metal Organic Frameworks (MOFs) with High CO 2 Working Capacity and CO 2 /CH 4 Selectivity for Methane Purification. *Eur. J. Inorg. Chem.* **2016,** 4505–4511 (2016).

132. Fernandez, M., Boyd, P. G., Daff, T. D., Aghaji, M. Z. & Woo, T. K. Rapid and Accurate Machine Learning Recognition of High Performing Metal Organic Frameworks for CO 2 Capture. *J. Phys. Chem. Lett.* **5,** 3056–3060 (2014).

133. Fernandez, M. & Barnard, A. S. Geometrical Properties Can Predict CO 2 and N 2 Adsorption Performance of Metal–Organic Frameworks (MOFs) at Low Pressure. *ACS Comb. Sci.* **18,** 243–252 (2016).

134. Fernandez, M., Trefiak, N. R. & Woo, T. K. Atomic Property Weighted Radial Distribution Functions Descriptors of Metal–Organic Frameworks for the Prediction of Gas Uptake Capacity. *J. Phys. Chem. C* **117,** 14095–14105 (2013).

135. Raccuglia, P. *et al.* Machine-learning-assisted materials discovery using failed experiments. *Nature* **533,** 73–76 (2016).

136. Evans, J. D. *et al.* Computational identification of organic porous molecular crystals. *CrystEngComm* **18,** 4133–4141 (2016).

137. Fernandez, M., Woo, T. K., Wilmer, C. E. & Snurr, R. Q. Large-scale Quantitative Structure-Property Relationship (QSPR) Analysis of Methane Storage in Metal-Organic Frameworks. *J. Phys. Chem. C* **117,** 7681–7689 (2013).

138. Lee, Y. *et al.* Pore-Geometry Recognition : On the Importance of Quantifying Similarity in Nanoporous Materials. *Nature communications* (2017). doi:10.1038/ncomms15396

139. Edelsbrunner, H. & Harer, J. *Computational Topology: An Introduction*. (American Mathematical Soc., 2010).

140. Boyd, P. G., Moosavi, S. M., Witman, M. & Smit, B. Force-Field Prediction of Materials Properties in Metal-Organic Frameworks. *J. Phys. Chem. Lett.* **8,** 357–363 (2017).

141. Springer, S. *et al.* A zeolitic imidazolate framework with conformational variety: conformational polymorphs versus frameworks with static conformational disorder. *CrystEngComm* **18,** 2477–2489 (2016).

142. Watanabe, T. & Sholl, D. S. Accelerating Applications of Metal–Organic Frameworks for Gas Adsorption and Separation by Computational Screening of Materials. *Langmuir* **28,** 14114–14128 (2012).

143. Nazarian, D., Camp, J. S., Chung, Y. G., Snurr, R. Q. & Sholl, D. S. Large-Scale Refinement of Metal−Organic Framework Structures Using Density Functional Theory. *Chem. Mater.* acs.chemmater.6b04226 (2016). doi:10.1021/acs.chemmater.6b04226

144. Witman, M. *et al.* The Influence of Intrinsic Framework Flexibility on Adsorption in Nanoporous Materials. *J. Am. Chem. Soc.* **139,** 5547–5557 (2017).

145. Krause, S. *et al.* A pressure-amplifying framework material with negative gas adsorption transitions. *Nature* **532,** 348–352 (2016).

146. Serre, C., Bourrelly, S., Ramsahye, N. A. & Maurin, G. An explanation for the very large breathing effect of a metal-organic framework during CO2 adsorption. *Adv. Mater.* **19,** 2246–2251 (2007).

147. Barthelet, K., Marrot, J. J., Riou, D., Férey, G. & Férey, G. A Breathing Hybrid Organic–Inorganic Solid with Very Large Pores and High Magnetic Characteristics. *Angew. Chemie Int. Ed.* **41,** 281 (2002).

148. Serre, C. *et al.* Very large breathing effect in the first nanoporous chromium(III)-based solids: MIL-53 or CrIII(OH)·{O2C-C6H4- CO2}·{HO2C-C6H4 -CO2H}x·H2Oy. *J. Am. Chem. Soc.* **124,** 13519–13526 (2002).

149. Barthelet, K., Marrot, J., Férey, G. & Riou, D. V III (OH){O 2 C–C 6 H 4 –CO 2 }.(HO 2 C–C 6 H 4 –CO 2 H)x(DMF) y (H 2 O)z (or MIL-68), a new vanadocarboxylate with a large pore hybrid topology : reticular synthesis with infinite inorganic building blocks? *Chem. Commun.* 520–521 (2004). doi:10.1039/B312589K

150. Tan, J. C. & Cheetham, A. K. Mechanical properties of hybrid inorganic–organic framework materials: establishing fundamental structure–property relationships. *Chem. Soc. Rev.* **40,** 1059 (2011).

151. Greathouse, J. A. & Allendorf, M. D. The Interaction of Water with MOF-5 Simulated by Molecular Dynamics. *J. Am. Chem. Soc.* **128,** 10678–10679 (2006).

152. Haigis, V., Coudert, F.-X., Vuilleumier, R., Boutin, A. & Fuchs, A. H. Hydrothermal Breakdown of Flexible Metal–Organic Frameworks: A Study by First-Principles Molecular Dynamics. *J. Phys. Chem. Lett.* **6,** 4365–4370 (2015).

153. Chanut, N. *et al.* Screening the Effect of Water Vapour on Gas Adsorption Performance: Application to CO 2 Capture from Flue Gas in Metal-Organic Frameworks. *ChemSusChem* **10,** 1543–1553 (2017).

154. Bellarosa, L., Gutiérrez-Sevillano, J. J., Calero, S. & López, N. How ligands improve the hydrothermal stability and affect the adsorption in the IRMOF family. *Phys. Chem. Chem. Phys.* **15,** 17696 (2013).

155. Vanduyfhuys, L. *et al.* QuickFF: A program for a quick and easy derivation of force fields for metal-organic frameworks from ab initio input. *J. Comput. Chem.* **36,** 1015–1027 (2015).

156. Wieme, J., Vanduyfhuys, L., Rogge, S. M. J., Waroquier, M. & Van Speybroeck, V. Exploring the Flexibility of MIL-47(V)-Type Materials Using Force Field Molecular Dynamics Simulations. *J. Phys. Chem. C* **120,** 14934–14947 (2016).

157. Bureekaew, S. *et al.* MOF-FF - A flexible first-principles derived force field for metal-organic frameworks. *Phys. status solidi* **250,** 1128–1141 (2013).

158. Bristow, J. K., Tiana, D. & Walsh, A. Transferable Force Field for Metal–Organic Frameworks from First-Principles: BTW-FF. *J. Chem. Theory Comput.* **10,** 4644–4652 (2014).

159. Bristow, J. K., Skelton, J. M., Svane, K. L., Walsh, A. & Gale, J. D. A general forcefield for accurate phonon properties of metal–organic frameworks. *Phys. Chem. Chem. Phys.* **18,** 29316–29329

(2016).